

Zur interkulturellen Validität von Meßinstrumenten

Wendt-Hildebrandt, Susan; Hildebrandt, Kai; Krebs, Dagmar

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Wendt-Hildebrandt, S., Hildebrandt, K., & Krebs, D. (1983). Zur interkulturellen Validität von Meßinstrumenten. *ZUMA Nachrichten*, 7(13), 45-57. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210489>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ZUR INTERKULTURELLEN VALIDITÄT VON MESSINSTRUMENTEN

1. Einführung: Das Gesamtprojekt "Interkulturelle Skalenentwicklung"

In diesem Projekt wird die Übertragung der in den USA entwickelten "social life feeling scales" in den deutschen Kontext angestrebt. Diese insgesamt 14 Skalen (12 allgemeine Einstellungsskalen und zwei Skalen zur Erfassung der sozialen Wünschbarkeit) müssen zu diesem Zweck zunächst sehr sorgfältig übersetzt und im Hinblick auf die Verständlichkeit der Übersetzung und die technische Anwendbarkeit der gegebenen Stimuli überprüft werden. In einem weiteren Schritt ist der interkulturelle Vergleich der Skalen mittels statistischer Verfahren geplant.

Ausgangspunkt für dieses Projekt war das Interesse an den Themenbereichen der Entfremdung und der zunehmend negativen und pessimistischen Weltsicht, dem sog. "no future"-Syndrom. Sowohl aus forschungsökonomischen Gründen als auch der Vergleichbarkeit wegen wurden zur Erfassung dieser allgemeinen Stimmungen nicht neue Skalen konstruiert, sondern es wurde auf die von SCHUESSLER (1982) veröffentlichten Skalen zurückgegriffen. Diese Skalen beziehen sich auf die in der individuellen Wahrnehmung repräsentierte Möglichkeit der aktiven Streuung und/oder Einflußnahme auf Prozesse und/oder Zustände im sozialen, beruflichen und politischen Bereich. Weiterhin beinhalten die Skalen allgemeine Aussagen über die gegenwärtige Gesamtstimmung bezüglich der gesellschaftlichen, aber auch der persönlichen Zukunft. Die Einzelfragen (Items) wurden von SCHUESSLER aus der anglo-amerikanischen Literatur zusammengetragen, explorativ faktorenanalytisch ausgewertet und zu den inhaltlichen Bereichen Selbstbestimmung, Vertrauenswürdigkeit der Mitmenschen, allgemein negative Stimmung, Jobzufriedenheit, politisches Engagement, allgemein positive Stimmung, Zynismus, politische Desillusioniertheit, Zukunftsperspektiven, ökonomische Selbstbestimmung, allgemeine Demoralisierung und Karriereorientierung zusammengefaßt.

Im April 1983 wurde bei ZUMA ein erster Pretest (im Gegensatz zur Pilot-Studie - siehe unten - nicht mit Studenten) durchgeführt mit dem Ziel, die Verständlichkeit der Items und mögliche Ungenauigkeiten in der Übersetzung - resultierend aus der Übersetzungsproblematik bei idiomatischen Redewendungen und kulturspezifischen Aussagen - herauszuarbeiten. In diesem Pretest wur-

den die Interviewer dazu angehalten, Kommentare und Rückfragen der Befragten wörtlich zu notieren. Diese Aufzeichnungen waren einerseits sehr nützlich bei der Umformulierung von Items, machten andererseits jedoch auch die Notwendigkeit einer Rückübersetzung der Items ins Amerikanische deutlich, deren Ziel es sein sollte zu überprüfen, inwiefern der Sinngehalt der Items durch die Übersetzung möglicherweise verändert worden war.

Die Itemformulierungen wurden entsprechend den Ergebnissen und Hinweisen des Pretests verändert und dann in einer Pilot-Studie Studenten der Universitäten Gießen, Hamburg und Wuppertal zur schriftlichen Beantwortung vorgelegt. Erste Ergebnisse dieser Pilot-Studie liegen vor, sind jedoch nicht Gegenstand dieses Berichts. Hier wollen wir neben allgemeinen Überlegungen zur Methode über konkrete Erfahrungen mit der Rückübersetzung der Skalenitems in diesem Projekt berichten.

2. Funktionale Äquivalenz von Übersetzungen und interkulturelle Validität von Meßinstrumenten

Bei vergleichenden Analysen, die Sprachgrenzen und damit natürlich auch Kultur- oder Systemgrenzen überschreiten, stellt sich die Frage der Äquivalenz automatisch, da sie das einzig sinnvolle Bewertungskriterium bei der notwendigen Übersetzung des Meßinstruments ist. Die Forderung lautet also, durch eine funktional äquivalente Übersetzung die Stimuli konstant zu halten und damit für die anzustellenden Vergleiche auszuschalten oder neutral zu halten.

Das Resultat einer Variablenmessung wird u. a. von der "wirklichen" Ausprägung des zu messenden Phänomens und von der Meßmethode bestimmt. "If it can be assumed that the measurement of a given variable is relatively free of systematic error at the system level ... a simple test ... can be used to ascertain whether social systems differ with regard to this variable" (PRZEWORSKI & TEUNE, 1970:40, unsere Hervorhebung). Solange der Einfluß der Meßmethode für alle Messungen konstant ist, kann man zwei Messungen direkt zu vergleichenden Aussagen über die zu messenden Phänomene selbst benutzen. Diese Annahme wird bei Untersuchungen, die einen Kulturkreis (politisches, soziales System) nicht verlassen, meist nicht problematisiert. Auch im selben Kulturkreis kann aber die Meßmethode durchaus "systematic errors at the

system level" hervorrufen, wenn nämlich die Angehörigen verschiedener Subsysteme unterschiedlich auf die Meßstimuli reagieren (VERBA et al., 1978:34).

Das Problem der funktionalen Äquivalenz von Übersetzungen fällt also in den generellen Bereich der Validitätsfragen (ANDERSON, 1967:126). Die Validitätsproblematik ist bei vergleichenden Analysen in besonderer Schärfe gestellt, weil es zum einen um die (relativ gleiche) Validität von zwei (oder mehr) Instrumenten in unterschiedlichen Kontexten geht, statt um ein einziges Instrument, und weil zum zweiten die üblichen Methoden, Validität zu prüfen, bei der Anwendung in unterschiedlichen Systemen komplizierter oder überhaupt fraglich werden (BLALOCK, 1982). Normalerweise kann die Annahme von Validität auf drei Arten überprüft werden, die kurz als "face validity", Konstruktvalidität (innere Gültigkeit) und Kriterienvalidität (äußere Gültigkeit) bezeichnet werden. Die beiden letzteren Gültigkeitstests sind aber bei vergleichenden Forschungen problematisch, weil etwa Art und Grad der Beziehungen zwischen der Untersuchungsvariablen und möglichen Kriterienvariablen oft ja gerade erst analysiert werden sollen. Die Gleichartigkeit dieser Beziehungen in verschiedenen Systemen kann dann nicht a priori angenommen werden. Kritischer noch ist, daß das Vorhandensein von Variablen, die selbst bereits valide sein müssen, Voraussetzung ist für ihre Benutzung in der Validierung. Sowohl bei den Kriterien- wie bei den Konstruktvariablen läuft man deshalb Gefahr, in eine unendliche Regression zu gelangen, es sei denn, man hat bereits unabhängig validierte Variablen; man braucht also Seiteninformationen oder "contextual information". (Siehe einerseits die komplizierten Methoden über "auxiliary measurement error theories", die BLALOCK vorschlägt, um die Vergleichbarkeit von Meßinstrumenten/Indikatoren zu überprüfen, und die Anmerkungen bei FREY (1970) und HYMES (1970) zur Einbeziehung des Kontextes.)

Gegenüber den beiden letzteren Aspekten der Gültigkeit tritt die Prüfung der "face validity" in der Regel als relativ primitiv, da von subjektiven Einschätzungen der "intuitiven Plausibilität" abhängig, theoretisch zurück. Während in der Forschungsrealität aber die Abschätzungen der "face validity" oft der einfachste und deshalb der einzige Schritt zur Validitätsprüfung bleibt, so erweist sie sich in der vergleichenden Forschung nicht nur als der am leichtesten durchzuführende Schritt, sondern auch der, der die

wenigsten a priori Annahmen erfordert. Wenn die Übersetzung "stimmt", dann sollte das auch heißen, daß die beiden Instrumente funktional äquivalent sind - eben das macht ja eine "gute" Übersetzung aus -, jedenfalls soweit der Übersetzer das beurteilen kann. Zudem gibt es mit der "blinden Rückübersetzung" ein nützliches Werkzeug, mit dem man funktional äquivalente Übersetzungen auch nicht garantieren, aber doch, durch wiederholtes Übersetzen, Fehler sukzessiv verringern kann.

Damit schafft die durch Rückübersetzung erhöhte "face validity" erst die Grundlage für kompliziertere Validitätsprüfungen auch statistischer Art, wie sie von PRZEWORSKI & TEUNE (1967) in Anlehnung an SCOTT (1968) vorgeschlagen werden. Obwohl im Vergleich zu diesen die "blinde" Rückübersetzung also sowohl ein probates wie relativ billiges und im Prinzip einfaches Mittel zur Überprüfung der Äquivalenz von Instrumenten in unterschiedlichen Systemen ist, wird diese Methode doch relativ selten strikt angewandt. Im Gegensatz zur berühmten "Civil Culture"-Studie (ALMOND & VERBA, 1963) haben selbst so große und/oder anspruchsvolle Projekte wie die "Eurobarometer" oder die "Political Action"-Studie auf die Methode verzichtet. Die Gründe liegen nicht zuletzt darin, daß häufig Personal mit den notwendigen Qualifikationen fehlt. "The translator needs to be bicultural, not merely bilingual" (CASAGRANDE, 1954:338).

Es muß allerdings angemerkt werden, daß die in der Literatur über Methoden des Vergleichs beschriebenen Vorgehensweisen bei der Rückübersetzung durchaus aufwendig werden können. So schlagen etwa BRISLIN et al. (1973:453) nicht nur vor, Informationen über die relative Häufigkeit von Wörtern bei der Wortwahl zu berücksichtigen, sondern die Übersetzungen auch an vier Gruppen von zweisprachigen (!) Befragten zu testen, die jeweils verschiedene Kombinationen von Original- und übersetzten Items beantworten sollen. Um Unterschiede zwischen diesen Gruppen statistisch absichern zu können, benötigte man so viele "Bilinguals", daß solche Vorgehen in der Regel unrealistisch sein dürften.

Im folgenden soll exemplarisch über die Methode der "blinden Rückübersetzung" berichtet werden, wie sie im Rahmen des Projekts "Interkulturelle Skalenentwicklung" zur Überprüfung der übertragenen Skalenitems angewandt wurde.

3. Beschreibung des Vorgehens

Es ist für die Forschungspraxis sicher bezeichnend, daß auch in diesem Projekt die Rückübersetzung erst in einem relativ späten Stadium eingesetzt wurde. Die prinzipielle Wünschbarkeit eines solchen Vorgehens war wohl früher erkannt worden, der erwartete Nutzen aber nicht genügend hoch angesetzt worden, als daß eine Rückübersetzung durch professionelle Übersetzer mach- und bezahlbar gewesen wäre.

Die deutsche Fassung der Skalen war mit Sachverstand im doppelten Sinne (sprachlich und fachlich) erstellt worden; sie wurde keineswegs als vorläufig angesehen. Die Items wurden unter Mitwirkung von geschulten Übersetzern und Sozialwissenschaftlern ins Deutsche übersetzt und z. T. auch von deutschsprechenden Personen mit Englisch als Muttersprache überarbeitet.

Als sich jedoch bei der Pilot-Studie für einige der 14 Skalen wesentlich niedrigere Alpha-Reliabilitätskoeffizienten als bei der amerikanischen Studie ergaben, wurde die Rückübersetzung in Angriff genommen. Diese geschah vollkommen blind, d. h. es lagen jeweils nur die deutschsprachigen Fassungen vor, nicht aber die amerikanischen Originale. Es ist allerdings bei Sozialwissenschaftlern natürlich nicht auszuschließen, daß englische Originalitems auch in der genauen Formulierung aus der Literatur bekannt sind. Es war jedoch unser Eindruck, daß dies nur bei wenigen Items der Fall war. [Die Übersetzung wurde zunächst von einem durch langen Nordamerikaaufenthalt zweisprachigen Deutschen vorgenommen, und dann von einer ebenfalls zweisprachigen amerikanischen Germanistin überprüft bzw. in Zweifelsfragen wiederholt.

Im nächsten Schritt, den ... Schritt, werden

Der nächste Schritt stellt das Herzstück des Rückübersetzungsverfahrens dar: Es wurden für jedes Item das amerikanische Original, die ursprüngliche deutsche Fassung und die Rückübersetzung(en) ins Englische auf einer Karteikarte zusammengefaßt, auf der auch etwaige Anmerkungen aus dem Pretestbericht eingetragen wurden. In mehreren Sitzungen wurden die Karten dann nacheinander durchgesehen, wobei die beiden englischen Fassungen miteinander verglichen und die deutsche Übersetzung ggf. geändert wurden. Dabei wurde versucht, Diskrepanzen zwischen den beiden englischen Versionen im Sinne der von HYMES (1970:326) als einzig akzeptabel betrachteten Reaktion

face validity
and validity
John S. Lee

ZUMA

auf Rückübersetzungen als "evidence of stylistic, semantic, and social factors operative in responses to the original material" zu verstehen, um im Licht dieser Faktoren die deutschen Fassungen zu verändern und (hoffentlich) zu verbessern. Im Rahmen dieses kurzen Berichts kann natürlich die theoretische Problematik von Rückübersetzungen und, noch grundlegender, der funktionalen Äquivalenz nicht im Detail erörtert werden. Es soll aber darauf hingewiesen werden, daß die Forderung nach funktionaler Äquivalenz nicht nur schwer einzulösen ist, sondern ihre Erfüllung oder Nicht-Erfüllung auch nur sehr grob abgeschätzt werden kann. (Zu solchen Meßproblemen siehe BRISLIN et al., 1973:51f.) Die Rückübersetzung kann in dieser Hinsicht lediglich Hinweise geben. Im Prinzip muß nämlich auch eine wörtliche Übereinstimmung von Original und Rückübersetzung nicht bedeuten, daß funktionelle Äquivalenz zwischen Deutsch und Englisch vorliegt, wenn nämlich gleiche Worte Inhalte bezeichnen, die in beiden Kulturen unterschiedlich "geladen" sind. Ein Beispiel dafür wären die Begriffe Links und Rechts, die in Europa und in Nordamerika nicht nur unterschiedlich relevant ("salient") sind, sondern (deshalb) auch tendenziell andere Bedeutungen haben (BARNES, KAASE et al., 1979). Ein Beispiel in ähnlicher Richtung aus unserer Arbeit sind "poor people" - "Arme Leute", da die Armutsdiskussion in der amerikanischen Politik einen wesentlich größeren Stellenwert hat als in der deutschen, wo - in teilweiser Vorwegnahme des Diskussionsergebnisses über die Ursachen - eher von "sozial benachteiligten oder sozial schwächeren Schichten" die Rede ist.

Der Begriff "funktionale Äquivalenz" weist ja bereits darauf hin, daß die Übersetzungen in Bezug zu einem Kontext stehen, der je nach Untersuchungsobjekt etwa politischer, sozialer, kultureller oder anderer Natur sein kann. Die Kontexte sind natürlich systemspezifisch. Deshalb kann es nicht nur vorkommen, daß bei der Übertragung eines Items in die andere Sprache (und damit Kultur) die Akzente verschoben werden müssen (siehe oben), sondern auch, daß sich gar kein äquivalenter Sinn findet. PRZEWORSKI & TEUNE (1970) fordern deshalb etwa, daß Indikatoren, die interkulturell vergleichend angewandt werden sollen, sowohl gemeinsame wie je systemspezifische Items enthalten sollten. Im gegenwärtigen Projekt können entsprechende Überlegungen aber erst angestellt werden, wenn die übertragenen Skalen empirisch überprüft worden sind.

Ergebnis der Analyse

Für die Auswertung der Rückübersetzungsmethode in diesem Projekt wurde jedes Item vercodet nach dem folgenden Schema, das auf dem Vergleich des englischen Originals mit der Rückübersetzung ins Englische basiert. Hierbei gab es mehrere Möglichkeiten (mit Beispiel):

1.1 Die englischen Fassungen (Original und Rückübersetzung) stimmten im Sinngehalt und auch Wort für Wort überein. BEISPIEL: Original: Most people can be trusted; Deutsch: Den meisten Menschen kann man trauen. Rückübersetzung: wie O.

1.2 Die englischen Fassungen stimmten zwar nicht wörtlich, aber doch im Sinn überein. Es sollte angemerkt werden, daß diese Kategorie keineswegs "schlechter" ist als 1.1, da es bei der Stimuliübertragung allein um funktionale Äquivalenz geht. BEISPIEL: O: What happens in life is largely a matter of chance; D: Was im Leben geschieht, hängt meistens vom Zufall ab; R: Much in life depends on chance.

In beiden obigen Fällen wurde die deutsche Fassung nicht verändert. Manchmal ergab sich allerdings als Resultat des intensiven "Daraufstarrens", daß uns die Formulierung des deutschen Items verbesserungsbedürftig und -fähig erschien:

1.3 Trotz Übereinstimmung in den beiden englischen Fassungen wurde das deutsche Item geändert. Diese Änderungen waren aber lediglich stilistischer Natur und sollten den Sinn nicht verändern. Solche Änderungen sind nicht direktes Ergebnis der Rückübersetzung, sondern wären bei jeder wiederholten Beschäftigung mit den Items zu erwarten.

Eine zweite Gruppe von Möglichkeiten resultierte, wenn der Vergleich von Original und Rückübersetzung Sinnunterschiede zeigte. Bei dieser Konstellation ergaben sich mehrere Erklärungsmöglichkeiten:

2.1 Die deutsche Fassung war eine ungenaue Wiedergabe des englischen Originals. BEISPIEL: O: People feel like they belong where I work; D: Da, wo ich arbeite, haben die Leute ein Gefühl der Zusammengehörigkeit; R: Where I work people have a sense of belonging together. Neue Fassung: Da, wo ich arbeite, haben die Leute das Gefühl der Zugehörigkeit.

- 2.2 (Als Sonderfall von 2.1) Die deutsche Fassung war eine ungenaue Wiedergabe des Originals, weil das Original idiomatische Redewendungen enthielt ("I seem to be marking time these days"), oder sich durch "false cognates" (stammgleiche Wörter mit unterschiedlicher Bedeutung) sinnverzerrende Übersetzungen angeboten hatten: "to contro!" ≠ "kontrollieren". BEISPIEL: O: In my opinion, this country is sick; D: Meiner Meinung nach ist in unserem Land vieles nicht in Ordnung; R: In my view there is a lot that is wrong with this country; N: Meiner Meinung nach ist dieses Land kaputt.
- 2.3 Die Rückübersetzung war ungenau; in diesem Fall ist es möglich, daß die deutsche Fassung sehr wohl gut war. Bei dem einmaligen Auftreten dieses Falles wurde die deutsche Fassung dennoch geändert - nicht aufgrund, sondern unabhängig vom Ergebnis der Rückübersetzung.

Die theoretisch auch mögliche Sonderform 2.4, daß nämlich die Rückübersetzung ungenau war, weil die deutsche Fassung idiomatisch war, trat in unserem Experiment nicht auf. (Es muß angemerkt werden, daß es menschlich schon verständlich wäre, wenn die Rückübersetzer selbst die Ungenauigkeit ihrer Arbeit nicht erkennen oder jedenfalls nicht zugeben wollten. Jedoch hat auch der dritte Codierer, der an der Rückübersetzung selbst kein persönliches Interesse hatte, keine weiteren Fälle von ungenauen Rückübersetzungen entdeckt.)

Neben den genannten Entscheidungen, die über 90 % aller Fälle (Fall hier gleich Item/Frage) umfaßten, gab es drei interessante Sonderformen:

- 3.1 Die Rückübersetzung legte einen Fehler bloß, der aber entstanden war, weil die "genaue" Übersetzung im Pretest auf Schwierigkeiten gestoßen war. Es konnte in solchen Fällen nicht nachgeprüft werden, ob die Pretestprobleme ein Hinweis auf fehlende funktionale Äquivalenz waren, oder ob das Originalitem auch im Englischen ähnliche Probleme hervorgerufen hatte. (Anzumerken ist, daß zwar die Daten der amerikanischen Studie, nicht aber Kommentare der Befragten vorlagen.)
- 3.2 "Bei näherem Hinsehen", d. h. bei der Rückübersetzung, wurden Probleme beim englischen Originalitem entdeckt, etwa weil es doppelte oder zweideutige Stimuli enthielt. Diese Ungenauigkeiten könnten u. U. dazu füh-

ren, auf ein solches Item zumindest bei der vergleichenden Forschung zu verzichten (BRISLIN et al., 1973), oder das Original zu verändern, weil die Übertragung durchaus anwendbar sein kann. Die Entdeckung solcher Fälle ist nicht verwunderlich, weil manchmal erst der Prozeß des Übersetzens etwaige logische, sprachliche oder inhaltliche Ungenauigkeiten herausstellt, die man im Original auch bei zweimaligem Lesen leicht übersehen kann. BEISPIEL: O: Many people in our society are lonely and unrelated to their fellow human beings; D: Viele Leute in unserer Gesellschaft sind einsam und haben keine Beziehung zu ihren Mitmenschen.

3.3 In diesen Fällen wurden die Schwierigkeiten, eine funktional äquivalente Übertragung zu finden, besonders deutlich, weil Schlüsselwörter nicht oder nur sehr umständlich zu übersetzen waren. Englische Wörter, die eine solche Codierung auslösten, waren u. a.: Friends, public officials, government. Eine solche Codierung muß natürlich nicht bedeuten, daß es kein funktionales Äquivalent gibt, nur war eine äquivalente Übertragung schwerer zu finden. Es schien lediglich sinnvoll, solche Fälle gesondert zu zählen, weil damit ein Hinweis auf die tatsächliche Häufigkeit eines der im Prinzip schwierigsten Probleme bei der Verpflanzung von Meßinstrumenten gegeben wird.

Bei Items, die mit den Kategorien 3.2 und 3.3 vercodet werden, sollten im Idealfall auch die Originalitems wieder zur Diskussion gestellt werden (BRISLIN et al., 1973:36f.). Damit sollen die Instrumente ihres ethnozentrischen Ursprungs (sowohl im Inhalt wie in der Form) entkleidet ("decentered") werden. Nachdem die hier zu übertragenden Skalen aber in den USA bereits erprobt sind, ist eine solche Zielsetzung in unserem Fall nur begrenzt möglich.

4. Ergebnisse

Die 110 "Fälle" (Items) wurden nach diesem Schema von drei Personen blind codiert. Genau genommen wurde die Hälfte der Items nur "halb-blind" codiert; eine getrennte Auszählung ergab aber, daß der Grad der Übereinstimmung zwischen den Codierern in beiden Fällen fast identisch war (60% gegen 63,6% volle Übereinstimmung aller drei Codierer).

Nachdem die drei Codierer unabhängig ihre Einstufungen nach den oben aufgeführten Kategorien vorgenommen hatten, wurde in einer gemeinsamen Sitzung ein Code für jeden Fall erarbeitet, da es bei der Beurteilung der Nützlichkeit von Rückübersetzungen ja auf eine einzige Beurteilung der Übereinstimmung/Nichtübereinstimmung von Original und Rückübersetzung ankommt. Es ergab sich jedoch, daß in 61,8 % aller Fälle ohnehin volle Einigkeit über die Verschlüsselung bestand, und in weiteren 13,6 % nur ein Bearbeiter von der Entscheidung abwich. Lediglich in 18,2 % der Fälle hatte jeder Codierer eine andere Kategorie gewählt, und in weiteren 6,3 % ergab sich in der Diskussion der Diskrepanzen ein völlig neuer Wert.

Tab. 1: Ergebnisse der Rückübersetzung

<u>Kategorie</u>	<u>N</u>	<u>%</u>
<u>1. Rückübersetzung kein Hinweis auf Fehler</u>		
1.1 Wörtliche und sinngemäße Übereinstimmung von Original und Rückübersetzung	21	53.6 { 19.1 20.9 13.6
1.2 Nur sinngemäße Übereinstimmung von Original und Rückübersetzung	23	
1.3 Stilistische Änderungen der deutschen Fassung (<u>nicht</u> auf Rückübersetzung bezogen)	15	
<u>2. Rückübersetzung gab Hinweis auf Fehler</u>		
2.1 Englisch - deutsche Übersetzung ungenau	32	37.3 { 29.1 8.2
2.2 Wie 2.1, aber erklärbar durch idiomatische Wendung oder "false cognate" im Original	9	
2.3 Deutsch - englische Rückübersetzung ungenau (Änderung der dt. Fassung nicht notwendig)	1	0.9
<u>3. Sonderfälle</u>		
3.1 Ursprüngliche deutsche Fassung nach Pretest geändert	1	0.9
3.2 Originalitem problematisch (doppelte oder zweideutige Stimuli)	4	3.6
3.3 Kulturspezifische Items	4	3.6
Gesamt	110	99.9

Tabelle 1 faßt das Ergebnis der Auszählung zusammen: Danach ergibt sich zumindest in einer knappen Mehrzahl (53,6 %) der Items, daß die ursprüngliche Fassung der deutschen Version der Prüfung durch die Rückübersetzung standhielt. Andererseits wurden in fast 40 % der Fälle aufgrund der Rückübersetzung Veränderungen vorgenommen! Diesen relativ hohen Anteil kann man nicht, wie man ja bei Wissenschaftlern vermuten könnte, auf die ohnehin vorhandene hohe generelle Neigung zum Kritisieren zurückführen, da die Codes 2.1, 2.2 und 2.3 nur dann gewählt wurden, wenn die Rückübersetzung tatsächlich nicht mit dem Original übereinstimmte. Ein Hinweis auf die ungefähre Anzahl der Änderungen, die sich auch bei einfachem "noch mal durchsehen", d. h. ohne blinde Rückübersetzung, hätten finden lassen, geben die 15 Fälle mit dem Code 1.3, die also stilistisch verändert wurden, ohne daß dafür eine Notwendigkeit aus der Rückübersetzung abzuleiten war. Überträgt man diesen Prozentsatz (15 aus 59 = 25,4 %) auf die Gesamtmenge der Items, so müßte man allein aus der erneuten Durchsicht 28 Änderungen erwarten, statt der tatsächlich vorgenommenen 57: 15 stilistische + 42 "echte" Änderungen (Codes 2.1 + 2.2).

Das Fazit muß also lauten, daß der durchlaufene Prozeß zwar "schmerzhaft", aber durchaus ergiebig war in dem Sinne, daß eine Anzahl von nach unserem Dafürhalten nicht funktional äquivalenten Items im Netz der Rückübersetzung hängen blieben. Damit ist jedoch über die geänderten Formulierungen noch nichts ausgesagt: Streng gesprochen müßte der Prozeß der iterativen Rückübersetzung natürlich nun auf die neu formulierten Sätze angewandt werden - eine Forderung, die umso schwieriger durchzuführen ist, als die zweisprachigen Teilnehmer bei der oben referierten Übersetzung nicht noch einmal eingesetzt werden können, da sie nicht länger "blind" sind.

Weiterhin ist zu beachten, daß die Entscheidung über die endgültige Formulierung der deutschen Items wohl nicht bei Personen liegen sollte, die fließend zweisprachig oder gar "bicultural" sind. Solche Personen sind aufgrund ihrer Sprachkenntnis und der Lebensläufe, auf denen die Sprachkenntnis beruht, sicher untypisch für den "durchschnittlichen Befragten". Deshalb wird in der Literatur darauf hingewiesen, daß die "Endredaktion" bei einsprachigen Individuen liegen sollte, allerdings in Zusammenarbeit mit den zweisprachigen Übersetzern.

5. Ausblick und weitere Perspektiven

Da die Pilot-Studie an einer Spezialpopulation (Studenten) durchgeführt wurde, ist es notwendig, die nun erarbeitete Fassung der Skalen noch einmal in einem Pretest zu erproben, bevor sie einer repräsentativen Stichprobe von Personen (geplant ist ein Sample von $N = 2000$) zur Beantwortung vorgelegt werden.

Im weiteren Verlauf des Projekts wird neben der Überprüfung der Dimensionen der verwendeten Konstrukte der Vergleich mit den Ergebnissen der amerikanischen Studie einen zentralen Stellenwert haben. Insbesondere wird auf die Herausarbeitung inhaltlicher Unterschiede in der Reaktion auf die Items und in der Struktur der Items Wert gelegt. Für die Beurteilung dieser in der Wirklichkeit bestehenden Unterschiede ist eben die Plausibilität der Annahme, daß die Meßinstrumente funktional äquivalent und damit ohne verzerrenden Einfluß sind, grundlegend wichtig. Die Rückübersetzung der Items hat hoffentlich dazu beigetragen, das Auftreten von solchen Artefakten, die sich durch Sinnverschiebungen in den Formulierungen ergeben können, zu reduzieren. Ob aber die Änderungen wirklich auch Verbesserungen darstellen, kann beim gegenwärtigen Stand des Projekts allenfalls im Sinne von verbesserter "face validity" behauptet werden. Ein echter Test für diese Behauptung - und damit für den Nutzen der Rückübersetzungsprozedur - kann nur durch die empirische Anwendung der Items geleistet werden. Dabei sollen die alten und die per Rückübersetzung verbesserten Versionen derselben Items Befragten vorgelegt und die Items in bezug auf die Häufigkeit von fehlenden Antworten ("missing data"), die Randverteilungen und besonders die Skalierbarkeit miteinander verglichen werden.

Dieser Bericht wurde verfaßt von Susan Wendt-Hildebrandt (Universität Windsor) und Kai Hildebrandt (ZUMA), die beide die Rückübersetzung der Items besorgten, sowie von Dagmar Krebs, die das Gesamtprojekt bei ZUMA betreut.

Literatur

- ALMOND, G. A. & VERBA, S. The civic culture. Princeton: Princeton University Press, 1963.
- ANDERSON, B. W. On the comparability of meaningful stimuli in cross-cultural research. Sociometry, 30, 1967, 119-135.
- BARNES, S. & KAASE, M. et al. Political Action: Mass Participation in 5 Western Democracies. Beverly Hills: Sage, 1979.
- BLALOCK, H. M. Conceptualization and measurement in the social sciences. Beverly Hills: Sage, 1982.
- BRISLIN, R. W., LONNER, W. J. & THORNDIKE, R. M. Cross-cultural research methods. New York: Wiley, 1973.
- CASAGRANDE, J. B. The ends of translation. International Journal of American Linguistics, 20, 1954, 336-343.
- FREY, F. W. Cross-cultural survey research in political science. In: R. T. HOLT & J. E. TURNER (Eds.). The methodology of comparative research. New York: Free Press, 1970, 173-294.
- HYMES, D. Linguistic aspects of comparative political research. In: R. T. HOLT & J. E. TURNER (Eds.). The methodology of comparative research. New York: Free Press, 1970, 295-342.
- PRZEWORSKI, A. & TEUNE, H. Equivalence in cross-national research. Public Opinion Quarterly, 30, 1967, 551-568.
- PRZEWORSKI, A. & TEUNE, H. The Logic of comparative social inquiry. New York: Wiley-Interscience, 1970.
- SCHUESSLER, K. Measuring social life feelings. San Francisco: Jossey-Bass, 1982.
- SCOTT, W. A. Attitude measurement. In: G. LINDZEY & E. ARONSON (Eds.). Handbook of social psychology. Cambridge: Addison-Wesley, rev. ed. 1968, 204-272.
- VERBA, S., NIE, N. H. & KIM, J. Participation and political equality. Cambridge: Cambridge University Press, 1978.